AFRL-RH-AZ-TR-2011-0007

Multi-INT and Information Operations Simulation and Training Technologies (MIISTT)

Michelle Caisse
L-3 Communications

Ronnie F. Silber
Lockheed Martin

Raymond T. Tillman
L-3 Communications

November 2010
Final Report

DISTRIBUTION A. Approved for public release; distribution unlimited. Approval given by 88ABW/PA, 88ABW-2011-2386, 29 Apr 11.

AIR FORCE RESEARCH LABORATORY
711TH HUMAN PERFORMANCE WING,
HUMAN EFFECTIVENESS DIRECTORATE,
MESA, AZ 85212
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

NOTICES

This report is published in the interest of scientific and technical information exchange and its publication does not constitute the Government's approval or disapproval of its idea or findings.

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

The Government's rights to use, modify, reproduce, release, perform, display, or disclose any technical data or computer software contained in this report are restricted by paragraph (b)(4) of the Rights in Noncommercial Technical Data and Computer Software, Small Business Innovation Research (SBIR) Program clause (DFARS 252.227-7018 (June 1995)) contained in the above identified contract. No restrictions apply after the expiration date shown above. Any reproduction of technical data, computer software, or portions thereof marked as SBIR data must also reproduce the markings.

Qualified requestors may obtain copies of this report from the Defense Technical Information Center (DTIC) at http://www.dtic.mil.

AFRL-RH-AZ-TR-2011-0007 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

OMAYRA GENAO Program Manager	HERBERT H. BELL Technical Advisor

JOEL D. BOSWELL, Lt Col, USAF
Chief, Warfighter Readiness Research Division
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* 30 Nov 2010	**2. REPORT TYPE** Final Report	**3. DATES COVERED** *(From - To)* 8/2006-11/2010

4. TITLE AND SUBTITLE
Multi-INT and Information Operations Simulation and Training Technologies (MIISTT) Final Report

5a. CONTRACT NUMBER
FA8650-05-D-6502

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Michelle Caisse, Ronnie F. Silber, Raymond T. Tillman

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
2830HXA1

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

L-3 Communications
6030 South Kent Street
Mesa, AZ 85212

Lockheed Martin
6030 South Kent Street
Mesa, AZ 85212

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Readiness Research Division
6030 South Kent Street
Mesa, AZ 85212-6061

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL;

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-RH-AZ-TR-2011-0007

12. DISTRIBUTION / AVAILABILITY STATEMENT
DISTRIBUTION A. Approved for public release; distribution unlimited. Approval given by 88 ABW/PA, 88ABW-2011-2386, 29 Apr 11.

13. SUPPLEMENTARY NOTES
Air Force Research Laboratory Program Manager: AFRL/RHAC, Chin K. Tam, 937-255-2045, DSN 785-6718

14. ABSTRACT

The Multi-INT and Information Operations Simulation and Training Technologies (MIISTT) research effort investigated ways to create synthetic speech output in multiple languages within the eXpert Common Immersive Theater Environment (XCITE), a Computer Generated Force (CGF) simulator. It applies Hidden Markov Model (HMM) speech synthesis and a database containing fixed and dynamic message components in multiple languages to generate realistic dynamic tactical radio communications. Communications are triggered by events that occur during a range of simulated tactics within the force-on-force simulation environment. This allows the creation of realistic dynamic training scenarios that include situation-dependent dynamic audio messaging for players of multiple international forces.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Chin K. Tam
a. REPORT UNCLASSIFIED	**b. ABSTRACT** UNCLASSIFIED	**c. THIS PAGE** UNCLASSIFIED	UNLIMITED	31	**19b. TELEPHONE NUMBER** *(include area code)* 937-255-2045

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

This page intentionally left blank.

ABSTRACT

The Multi-INT and Information Operations Simulation and Training Technologies (MIISTT) research effort investigated ways to create synthetic speech output in multiple languages within the eXpert Common Immersive Theater Environment (XCITE), a Computer Generated Force (CGF) simulator. It applies Hidden Markov Model (HMM) speech synthesis and a database containing fixed and dynamic message components in multiple languages to generate realistic dynamic tactical radio communications. Communications are triggered by events that occur during a range of simulated tactics within the force-on-force simulation environment. This allows the creation of realistic dynamic training scenarios that include situation-dependent dynamic audio messaging for players of multiple international forces.

Table of Contents

Multi-INT and Information Operations Simulation and Training Technologies (MIISTT)-Final Report

1.0 SCOPE OF WORK AND OBJECTIVES

The Warfighter Readiness Science & Technology Program Task Order 0026 Multi-INT and Information Operations Simulation and Training Technologies (MIISTT) covered the period of August 2006 to November 2010. The Statement of Work (SOW) paragraphs 1, 6.1, and 6.2 describe the scope of effort for research and development conducted by this task order during this period.

> 1. SCOPE. The objective of this effort is to develop realistic training environments for the ISR and IO training and readiness domain by developing advanced Multi-INT and Information Operations simulation and training Technologies (MIISTT) including realistic training environments, competency based scenarios, and training and performance assessment methodologies.
>
> 6.1 Audio Database Development. The contractor shall research, develop, and validate an advanced audio database for the initial foreign language being implemented in Task Order 0007 of this contract. The contractor shall collaborate with the AFRL/HEC to research, develop, and validate advanced audio databases in additional foreign languages to be implemented into the RFTE. Total number of languages developed shall be dependent upon available resources and success in implementing the initial foreign language. The contractor shall verify the contents of the audio databases with Subject Matter Experts to determine validity.
>
> 6.2 Software Development. The contractor shall develop sharable audio-based software and any additional software necessary to implement the RFTE into a realistic training environment. Software shall be capable of using digitized voice and audio in English and foreign languages which can be manipulated and/or synthesized to generate a dynamic operational scenario. The suite shall maintain the capability of rapidly developing dynamic training scenarios and incorporating them into DMO training.

1.1 Overview

The work performed under this contact continued earlier work, adding realistic and context-appropriate multi-language audio and text messaging output to an existing suite of software. The main components of the existing suite are:

- eXpert Common Immersive Theater Environment (XCITE), a Computer Generated Force (CGF) simulator engine
- ThreatIOS, an operator user interface to XCITE for building and running scenarios

The messaging portion of the product suite is called Radio Frequency Threat Environment (RFTE). Additional system capabilities developed or augmented during the performance of this effort are described in this report.

1.2 Specific Objectives

In order to accomplish the objectives of the SOW, work focused on five major areas:

1. Selection and integration of a speech synthesis technology
2. Building a message database
3. Developing voices
4. Developing code for message generation in XCITE
5. Developing voice server software to integrate and deliver synthetic speech to XCITE

2.0 BACKGROUND AND HISTORICAL CONTEXT

Two earlier task orders, TO 0017 and TO 0007, researched the feasibility of adding speech output to XCITE (Silber & Burlant, 2005; Tygret, Silber, Burlant, & Hoefer, 2005). Those efforts successfully integrated both digitized natural speech and synthetic speech and implemented one foreign language.

In November 2006, the TO 0026 effort began and was exclusively devoted to the development of a multi-language ISR training environment. In April 2007, work on the language database, voice development, and the speech synthesizer was transferred from TO 0007 to TO 0026, with TO 0007 then focusing on Command, Control, Communications, Computers, Intelligence, Surveillance and Reconnaissance (C4ISR) systems development and integration.

The contractor team at the Air Force Research Laboratory, 711th Human Performance Wing, Human Effectiveness Directorate, Warfighter Readiness Research Division (711 HPW/RHA) was supported by the 711th Human Performance Wing, Human Effectiveness Directorate, Anticipate & Influence Behavior Division (711 HPW/RHX) speech research team on matters relating to speech synthesis technologies and voice generation.

During the period of performance, the team also worked on a commercial contract for the same technology *RFTE Engineering Development, Integration & Delivery Support* for L-3 Communications Integrated Systems (L3/IS). The commercial contract customer is one of the TO 0026 customers. Therefore, the features and deliverables implemented for the commercial contract needed to be maintained for the final delivery of TO 0026.

2.1 MIISTT Team

Contractors

- Kyle Tygret (Systems Engineer and Functional Area Lead 11/2006 – 11/2009)
- Tom Knapp (Systems Engineer and Functional Area Lead 11/2009 – present)
- Ronnie Silber (Applied Linguist 11/2006 – present)
- Ray Tillman (Software Engineer 04/2007 – present)
- Rob Creter (Hardware Engineer 03/2008 – present)
- Jim Burlant (Software Engineer 11/2006 – 02/2009)
- Sam Hoefer (Software Engineer 11/2006 – 10/2009; Task Order Lead 09/2007 – 10/2009)
- Darrin Woudstra (Intern 6/2008 – 08/2008; Computational Linguist 01/2009 – present)
- Michelle Caisse (Computational Linguist 01/2009 – present; Task Order Lead 05/2010 – present)

Government Principle Investigator (PI)

- 08/2008 – 06/2009 Geoff Barbier
    - Lt. Matthew Linford, Alternate
- 07/2009 – 04/2010 Lt Omayra Genao
- 04/2010 – end Lt Chin Ki Tam

2.2 Message Database Development

The MIISTT project required a set of messages appropriate for common and significant military scenarios in all languages implemented. Subject matter experts (SMEs) were needed to provide the domain-specific messages. The following timeline documents the efforts to obtain message sets.

04 – 08/2007 Received assistance from Major Guevara to find sources of voice and message data for MIISTT.

08 – 12/2007 Received engineering research on ground order of battle and implementation of tank, artillery, forward observer, and command/control database development.

01/2008 Completed effort to insert new ground force communication messages into the RFTE message database.

02/2008 Completed updating existing radio communication models and adding new models in the RFTE database according to the current list of platforms available in XCITE.

02/20/2008 Completed a comprehensive list of existing message texts from the RFTE database, to be used by SMEs to provide message inputs in other languages.

04-09/2008 Worked with SMEs on site, at Offutt Air Force Base (AFB), and by email to collect information on airmen positions, source material and scenario development and develop message texts in a foreign language.

10/2008 Began data extraction from SME provided materials into the message tables.

02-04/2009 Initiated message database development in additional languages. Traveled to Offutt AFB the week of April 6th to meet with SME contacts in multiple languages for the purpose of information gathering. Continued contact by email and received additional database inputs.

05 – 06/2009 Populated the Tactical Message Database for additional languages and identified areas where additional software development is required.

11/2009 – 08/2010 Worked with Government PI and Patch Plus Consulting to try to obtain additional SME input for foreign language scenario development. Provided spreadsheet of desired data for language scenario development to Government PI. With Government PI, explored the potential of accessing domain specific language SMEs from the Davis Monthan Compass Call resources. Identified a potential contact within NSA who may be able to assist with SME input.

08 – 09/2010 SMEs from Offutt AFB visited the 711 HPW/RHA to provide support for the message database in multiple languages.

2.3 Voice Model Development

For each language implemented, a set of distinct synthetic voices was required. The following timeline documents the process of developing and obtaining voice models.

03 – 04/2007 Received downloaded HMM-based Speech Synthesis System (HTS) voices from the Warfighter Interface Division, Collaborative Interfaces Branch (711 HPW/RHCP); 711 HPW/RHA contract engineers tweaked parameters per domain.

03 – 07/2007 Researched data sources for Limited Domain (LDOM) synthesis of additional voices and languages: Linguistic Data Consortium, Army Research Laboratory / West Point Military Academy

04 – 10/2007 Determined that it was necessary to record audio source data for LDOM synthesis. Investigated obtaining audio recording hardware and small recording studio. Location of a small recording studio was approved and security approval for electronic equipment. Furniture and RFTE workstations were relocated to accommodate the recording studio. Voice recording computer software was installed and System Security Authorization Agreement additions for computer installation were coordinated.

07/2007 – 05/2008 Purchased Cepstral foreign language voices. Pursued acquisition of license for its use. Integrated with XCITE.

01 – 02/2008 Established a collaborative working arrangement with 711 HPW/RHX for the voice development effort. Telephone conference held on January 25, 2008 to discuss potential roles and responsibilities. 711 HPW/RHX provided guidance on a new voice generation process and tools for a LDOM HTS voice.

01/2008 Continued to build English HTS voice to validate development procedure.

02/2008 711 HPW/RHX provided an example of the new voice generation process and tools for a LDOM HTS voice.

03/2008 Received first custom HTS voices from 711 HPW/RHX.

03/2008 Received guidance on foreign language selection.

05/2008 SME visited 711 HPW/RHX to perform quality check on several text-to-speech (TTS) voices. Effort resulted in 26 unique TTS voice models to be delivered to AFRL/RHA.

06/2008 711 HPW/RHX delivered 26 unique voice models. Voice models were integrated into the HTS Voice Server.

06 – 07/2008 Resolved licensing issues for the foreign language source data used to generate HTS voice models. Two commercial licenses are required to gain access to and use the source data. Submitted purchase requests for commercial data licenses from XLingual (GlobalPhone) and Linguistics Data Consortium. Prepared a sole source letter in place for GlobalPhone license purchase request. This license is necessary for 711 HPW/RHA to receive a foreign language voice model from 711 HPW/RHX.

08/2008 711 HPW/RHA intern completed the HTS voice development procedure and held a training session with the team of engineers.

09/2008 Received all necessary data licenses from XLingual and Linguistic Data Consortium.

09/2008 – Received foreign language voice models from 711 HPW/RHX and integrated

07/2009 them into the HTS Voice Server.

09/2008 Created four LDOM TTS Voices for one language.

01/2009 Integrated updates to initial voice model files from AFRL/RHX.

01/2009 SME visited 711 HPW/RHX for voice model evaluation in a foreign language, resulting in the generation of multiple voice models for foreign languages.

03/2009 Received new licenses from Linguistic Data Consortium which allowed the use of new and improved voice models from 711 HPW/RHX. The Linguistic Data Consortium supports language-related education, research and technology development by creating and sharing linguistic resources: data, tools, and standards.

2.4 Software Development

Development work was required to add messaging capability to XCITE, integrate synthesis engines, expand Radio Frequency (RF) Host features, and develop a Voice Server.

06/2007 Integrated HTS synthesizer and Festival for client/server application.

09 – 12/2007 RF Host system software corrections, test, and integration. RF Host software was configured in Concurrent Versions Source (CVS) configuration management to assist in multi-user software development.

09/2007 – 03/2008 RFTE demonstration synthetic environment and XCITE synthetic environment software was merged for XCITE capabilities in ground order of battle.

10/2007 Integrated Cepstral voice data into audio processes.

03/2008 Successfully executed the HTS text to speech engine delivered from 711 HPW/RHCP.

04/2008 Began work on an HTS Engine Server to replace the current Festival Voice Server. Completed Word Dictionary Lookup application as part of the HTS Voice Server.

04/2008 Began updating the existing RF Host application to incorporate improvements from other programs.

08/2008 HTS Voice Server application was completed and work was begun on the DIS-enabled Voice Server monitor application.

09/2008 Initial L1 voice delivered to support integration of that language model into the HTS Voice Server application.

09/2008 RF Host/Simulation Environment Generation User Interface (SEGUE) Java Messaging Service interface was completed and tested using a test application provided by L3/IS, Greenville, TX.

10/2008 Voice Server monitor application which provides status of the HTS Voice Server to RF Host was completed.

11/2008 Delivered the Interim RFTE System to L3/IS in Greenville, TX.

04/2009 Implemented new noise mixing algorithms that allow the use of different platform and radio noise files as well as a user option to set signal to noise ratio.

05/2009 Worked with SMEs to create custom text transliteration schemes in multiple languages.

05/2009 Final RFTE software delivery performed for languages available to date.

07/2009 Completed beta version of the tactical database tool to be used for managing SME database input. Began testing tactical database tool by entering data content for several languages.

07/2009 Redesigned of message algorithms in the XCITE software. Architectural changes were required as the software scope expanded to include additional languages.

11/2009 – 08/2010 Added capabilities to RFTE Voice Server to accommodate grammar requirements.

11/2009 Created a small client of the Voice Server to allow testing of language generation by entering text rather than being driven only from XCITE scenarios.

01/2010 Created new branches in XCITE configuration management system to support control for Voice Server and RF Host.

02 – 11/2010 Tested Voice Server and XCITE/RFTE for final work with SMEs and product delivery. Fixed bugs as found.

2.5 **SME Visits**

SMEs were needed to provide message sets for languages and to validate and verify the adequacy of voice models and linguistic modeling.

05/2008 SME traveled to 711 HPW/RHX on May 22-23 to evaluate TTS voice models and finalize voice models to be delivered to 711 HPW/RHA.

06/2008 SME visited 711 HPW/RHA from June 16-19 to develop message texts in a foreign language.

08/2008 SME visited on August 19^{-}20 to collect additional information on airmen positions, source material and scenario development.

01/2009 SME visited 711 HPW/RHX on January 14th for voice model evaluation in a foreign language.

02/2009 Two SMEs visited 711 HPW/RHA from February 16-19 to validate the message database entries.

2.6 **Travel**

01/2007 Sam Hoefer supported Joint Expeditionary Force Experiment 2008/Government PI at Hurlburt Field for RFTE and Deployable Ground Station (DGSx) integration.

11/2007 Sam Hoefer traveled to Greenville, TX to attend RFTE Technical Interchange Meeting at 645th Aeronautical Systems Squadron.

05/2008 711 HPW/RHA TO 0026 team traveled to Offutt AFB May 13-14 to speak with SMEs and observe 711 HPW/RHCP's demonstration of latest HTS voice synthesis.

07/2008 SME visited Offutt AFB the week of July 21st for information gathering.

04/2009 711 HPW/RHA TO 0026 team traveled to Offutt AFB to meet with SMEs in multiple languages for information gathering.

2.7 **Workspace and Hardware**

06/2007 MIISTT development was relocated to C4ISR testbed.

09/2007 Recording studio was installed. Furniture and computer systems configuration changed.

12/2007 RFTE development area was configured.

07/2008 Moved development equipment to new Intelligence, Surveillance, and Reconnaissance (ISR) testbed area.

No work was performed by contractors in August, September, and October of 2009 when 2009 funding expired and 2010 funding was not yet available.

3.0 MATERIALS

3.1 Purchased Books

After extensive research for military terms dictionaries to support the research program the following were purchased from private book dealers and online. This collection will be transferred to the government at the close of contract.

1) Al-Salloom, Yusif Bin Ibrahim. (January 1998). Military Terms Dictionary: Arabic-English. Maktaba Al 'Oubecan.
2) Bourla, Y. (1998). Dictionary of Military Terms. Israel: Dvir Publishing House.
3) Galimberti Jarman, B., Russell, R., Carvajal, C. S., & Horwood J. (Eds.) (1994). The Oxford Spanish Dictionary: Spanish-English/English-Spanish, 3rd Ed. New York, NY: Oxford University Press.
4) Gongzhao, LI. (2006). An English-Chinese Military Dictionary. Shanghai: Shanghai Foreign Language Education Press, 1st Edition. ISBN 9787810955232: AU$179.95:
5) John Butt, J., & Benjamin, C. (2004). A New Reference Grammar of Modern Spanish 4th Ed. Republic of Malta: McGraw-Hill.
6) Kayyali, S. M. (March 1994). Modern Military Dictionary: English-Arabic/Arabic-English. Hippocrene Books.
7) Kostrov, A. D. (2000). Russian-English, English-Russian Military Dictionary: 50000Terms. Minsk Kiev: Technical Dictionaries; Bilingual edition.
8) Qi, J., & Wu, J. (2006). *Ying Han C4ISR ji shu ci dian : English-Chinese C4ISR technical terms dictionary*. Beijing Shi: Guo fang gong ye chu ban she.
9) Schultz, T., & Kirchhoff, K. (Eds.) (May 5, 2006). Multilingual Speech Processing. Boston, MA: Elsevier Academic Press.

3.2 Hardware

The following hardware items were purchased under TO 0026:

- 1 CLH International, Inc. dual Xeon system
- 5 CLH International, Inc. Intel Pentium 4 Extreme x6800 systems
- Miscellaneous hardware including monitors, switches, disks, UPS, cables, etc.
- 5' x 5' sound isolation booth with associated hardware and supplies

3.3 Software

- Cepstral, LLC voices, software development kit, and distribution license
- XLingual GlobalPhone phonetic dictionary research licenses and corpora

3.4 Training

- Fundamentals of Information Systems Security training class

4.0 PROCEDURE

4.1 Technology Selection

During the period of performance of TO 0007 and into the early portions of TO 0026, LDOM synthesis was selected as the synthesis technology to be used for multilingual synthesis. LDOM synthesis is produced by recording a speech corpus that consists of only the words that will be used in a given application. These target words are recorded in phrases or sentences that provide a fairly neutral acoustic context. Alternatively, they may be recorded in contexts that cover all of the contexts that the application is expected to exploit. The first path is typically chosen when the recorded corpus is large and contains sufficient acoustic and phonetic variation to cover the relevant contexts expected in the application. The second path is chosen when the recorded corpus is small and there is also a desire to capture the required contexts. In either case, the speech that is recorded is spoken in a fairly monotonic voice. During the recording phase, each word that is to be used for synthesis is recorded multiple times in multiple phrases, in order to capture a sufficient number of target tokens and language variation (minimally in unit initial, medial and final position).

After the speech is digitally recorded, it is cut into linguistic units of desired size. These linguistic units typically range in size from half syllables to phrases. With smaller units there is greater flexibility when the units are pasted back together to form new utterances. With larger units, naturalness is preserved. When speech is synthesized, the best candidate unit for a particular context is selected from the multiple instances of the unit available in the recorded corpus.

This technology provides a high degree of naturalness under ideal synthesis conditions. Furthermore, the available research showed that it was ideal for data sets that were finite and determinate, as with military brevity code. Indeed, when LDOM synthesis works well it is second only to recorded spoken speech for naturalness. However, when it fails, it can be very poor. The most common problem, especially with a small corpus, is that the unit required may not exist or exists only in a context very different from the one required. In such cases there will be noticeable discontinuities in the speech. There are ways of handling missing units, but they may introduce their own problems.

Ultimately LDOM was abandoned in favor of Hidden Markov Model (HMM) synthesis. In HMM synthesis, rather than storing and recombining digitized speech units, statistical parametric models of speech are stored. At runtime, the models are concatenated and the algorithms which created the models of the waveform are then reversed to recreate it. In addition to improved quality of the synthesis, there were other benefits to using HMM synthesis:

1. The voices could be built from either unclassified recordings by SMEs or other speakers or from publicly available speech corpora.
2. New voices and new message data could be added at any time.
3. The development of voice models and message data could be decoupled, allowing the voices to be developed in an unclassified environment, while the message data was developed in a classified environment.
4. In principal, parameters such as overall pitch, speech rate, and loudness, can be modified prior to playback to create new variants, such as intonation variation and levels of emotion.

The ability to decouple voice model and message data development enabled the expansion and distribution of available team resources without requiring extra clearances. People without high level clearances could work on voice and speech development, while people with the requisite clearances but lacking computational linguistics knowledge could develop message databases.

4.2 Voice Model Development

Early in 2008 the team decided to use HMM synthesis for the project. We adopted the standard development toolkit then available, the Hidden Markov Model Toolkit (HTK) to produce the voice models and HMM-based Speech Synthesis System (HTS) for synthesizing speech from the voice models. During the summer of 2008, a summer intern developed an English voice using HMM synthesis and documented the process. He developed both triphone and context-dependent pentaphone synthesis models. The pentaphone models sounded more natural, probably due to better representation of prosodic phenomena such as pitch.

Following this effort, voice development by 711 HPW/RHX developed HTS triphone model voices from language corpora purchased from the Linguistic Data Consortium (LDC) and XLingual.

4.3 Message Database Development

The project required a set of messages that reflected realistic dialogs that would occur in the types of scenarios modeled by XCITE in the languages selected for implementation. Messages were first defined for English. A spreadsheet was developed listing the English messages by category. In this format, SMEs could provide equivalent messages in other languages.

Early plans called for monthly SME visits to collect and verify tactical database entries in multiple languages. The actual availability of SMEs fell far short of that goal. In the summer of 2008, SMEs made several visits to 711 HPW/RHA and the TO 0026 team also visited Offutt AFB to work with the SMEs. In February and April of 2009 there were several visits between 711 HPW/RHA and SMEs from Offutt AFB. In August and September of 2010, SMEs from Offutt AFB visited 711 HPW/RHA. On each occasion, SMEs provided messages in foreign languages that covered the subject areas of the English message spreadsheet. Initially, SMEs entered messages into the spreadsheets. Later, theywere manually inserted into the database. During their last visit in 2010, SMEs used the Microsoft Access data entry forms, constraining them to provide messages that corresponded one-to-one with existing English messages. If no English message corresponded to the target language message, the data was entered in a spreadsheet.

4.4 Software Development

Modifications were made to a number of software components to fulfill the requirements of TO 0026. These are discussed in the following sections.

RF Host

The RF Host was re-architected to be more generic and portable to a variety of speech technologies. Under the auspices of the L3/IS RFTE contract, the capability to interface with the Rivet Joint Training System (RJTS) was added.

XCITE

Extensive modifications were made to XCITE to incorporate a communications model and to add code that triggers messages at appropriate points in the simulation. RFTE comprises approximately 17,000 lines of code within XCITE, plus additional code to trigger messages interspersed throughout.

Access database interfaces were built to allow entry and assignment of voices. The message database was redesigned for more efficient message storage and to allow role-dependent messages. In conjunction with this change, the Access message form was also updated.

In 2009, the capability to overlay noise representing radio and platform background noise was implemented to provide a more realistic simulation of the target speech.

Voice Server

The Voice Server applies a number of processing steps to the input text string:

1. Conversion to upper case
2. Separation of digits from alphabetic strings
3. Handling of specially marked pronunciation strings
4. Application of grammatical rules
5. Conversion of numbers to words
6. Conversion of words to phonemes
7. Conversion of phonemes to triphones

Steps 4 and 5 involve language-specific code and Steps 4, 5, and 6 are partially data driven. Each language has its own pronunciation look-up dictionary and text to phoneme rules. There are also grammar look-up lists for some languages.

The integration of each new language required:

- adding the dictionary and voice models to the appropriate directories
- creating new text to phoneme rule files
- coding new number processing rules
- adding dictionary entries for numbers
- adding a grammar dictionary, if required
- verifying that the phoneme symbols used in the dictionary are the same as those in the voice model

ThreatIOS

Several interface items were added to the ThreatIOS to allow the user to assign a voice to a player, assign a proficiency level, and to send a message. ThreatIOS was also modified to permit the presentation of the text used for speech synthesis to be displayed in close proximity to the speaking player.

5.0 POST MORTEM

5.1 Scope

Initial planning for this project revealed that the customer has a need for a number of languages and dialects. Given the lack of availability of SMEs, time and attention required to develop each language, and a brief lapse in funding, it was not possible to complete all of the languages required.

It was also initially considered desirable to model emotion levels as they would normally be heard in the field. Since the synthesis of emotion is in its infancy and the acoustic and perceptual correlates of emotions are not well understood, this requirement was given a lower priority in favor of focusing on the synthesis of the messages themselves.

The following are required to fully implement a message:

- The message text
- A **#define** preprocessor statement in the code that provides a name and unique number to represent the message
- A specification of the conditions under which the message gets spoken
- A block of code that implements the tactic or other context in which the message is spoken
- Lines of code within that tactic or context that trigger the message (or message sequence) under the appropriate conditions

This means that linguistic, tactical, and programming skills are all required, as well as a good knowledge of XCITE code. The problem becomes even more complex because languages differ in when and under what conditions messages are spoken and how long a sequence of messages is associated with a given tactic. Managing the timing and interplay of these skills and finding available resources was a challenge. We have collected messages for which there is code for the context, but not for the triggers, and others for which there is not yet code that implements the tactic during which the message would be spoken.

5.2 **Resource Requirements**

As previously mentioned, this project required a variety of different skill sets:

- software engineering
- military linguist SME
- military tactics expertise
- computational linguistics
- descriptive linguistics

The strengths of available resources did not match project requirements at various times.

Military linguists are in short supply. There was difficulty obtaining their time, in spite of the fact that their skills were crucial for this project. They undergo lengthy training (Powers, n.d.), including military basic training and up to 68 weeks at the Language Defense Institute. It is not possible to train contractor personnel to the necessary level of proficiency in multiple languages during the course of this or similar projects. There simply is no substitute or work-around for lack of SMEs. Their unavailability resulted in much lost time.

In early 2009, it was decided that voice models would be developed by 711 HPW/RHX, where strong computational linguistic skill already existed. At 711 HPW/RHA, development of the voice server, dictionaries, test-to-phoneme rules, language modeling, message database, and XCITE programming continued.

5.3 **Cultural Differences in Messaging and Tactics**

Initial modeling of the message database and tactics was done on English. Most of the coding and message database structure was complete before other languages and cultures were encountered. There are features of the other languages that do not fit this structure well and are currently difficult to implement. They may:

- use a more specific word to refer to some object in the situational context

- use messages tied to different aspects of the context or activities in the situation

5.4 Entering Message Data

Early in the project, message data collected from SMEs was entered into spreadsheets organized by subject area. The SMEs found this easy because they are trained to think about their language in terms of scenarios that involved a sequence of messages. However, the message database is organized by individual messages and the sequences are only visible in the code that triggers the messages. Therefore, there were difficulties when contractor personnel entered the messages into the database. The task required choosing a message from the spreadsheet to match a message in the database that was presented out of context. To make the selection, one had to look at the code, or later, at documentation that presented some of the messages in context. Then, a corresponding message in the target language had to be located. This was awkward and time consuming.

At the end of the project, SMEs were entering data directly into the message database using a form. Because the SMEs had greater familiarity with the messages and context this worked better. However, there are cases in which the foreign language material simply does not match the English. In these cases, comments were entered in the database or the material was entered in a spreadsheet.

5.5 Use of Multiple Languages

It is not uncommon for pilots from other countries to use English words and units of measure. We can model the use of English words by using the native language phoneme set in the pronunciation of English. But we currently have no way to use both English and metric units in association with dynamically inserted values for the same simulated entity.

5.6 Product Usability

RFTE is complex software. It requires some basic system administration skills to install and configure, as well as specialized skills to design, develop, and run a scenario. These skills are available for pilot training, but need to be augmented for linguist training.

In order to take advantage of the ability to transmit new messages during a scenario using the Communications, Navigation, and Identification (CNI) Messages dialog, an RFTE operator needs to be a linguist and needs to be aware of any idiosyncrasies of the text input requirements, such as use of accent marks.

Inevitably there will be language issues discovered during testing by users that require changes to messages, dictionary entries, or grammar files. This will require an understanding of how all language components work together, as well as ability to use the text input, phoneme symbols, grammatical markup, and dynamic insert symbols.

6.0 RECOMMENDATIONS

6.1 Message Database User Interface

Collecting message sets for scenario creation was one of the most important tasks during this project. It turned out to be difficult both because of the difficulty of obtaining SME time and because of the complexity of integrating messages and associated information with XCITE program functionality.

The message database input form currently used to collect information from SMEs has multiple problems:

- It requires that SMEs provide information they do not have, such as message type.
- It does not provide information that SMEs need to correctly enter messages.
- It does not have fields for information that the product team needs from SMEs.
- It does not allow SMEs to enter data in a natural way.

SMEs must select message type (command, response, status, track update) from a combo box before entering a message. All messages are categorized by type. In many cases, the type is obvious, but in some cases it is counter-intuitive. The XCITE messaging system should be changed to remove these concepts.

The checkboxes labeled Tcmds, Scmds, and Trkcmds have no meaning for SMEs and may no longer be needed except in a few cases. The use of these features and checkboxes should be reviewed and removed if possible.

SMEs must select a trigger, which is represented in the form by either a brief description or a code name (#define name) such as “SM_PERM2LAND”. It is not always easy to intuit the meaning of the trigger from this minimal information. Furthermore, the needed trigger may not be available if the message to be inserted is a new one. The SME would have a difficult time knowing whether to use one of the existing triggers.

In order to enter messages in a natural way, the SME should be able to enter messages by scenario, that is, as a sequence of messages in a dialog that is tied to a particular tactic or situational context. A scenario-based organization is also necessary for the SME to be able to properly review and verify existing messages. A trigger point would have to be assigned, representing a point in the scenario that triggers the sequence of messages. To be spoken correctly, the messages must be sequenced properly and must be spoken by the appropriate person in the scenario. Currently this information is in the code. Storing this information in the database in the form of sequence number and controller/controllee role would make a more rational interface between code and database, and would markedly simplify the process of collecting message data from SMEs.

A scenario-based organization of messages would allow different languages to have different numbers of messages in a scenario without cumbersome hard-coding of message sequences. It would enable SMEs to enter new messages to the message set in some cases without new coding. Some thought would need to be given to how trigger points are described to the user and how to collect information about new trigger points required by new message sequences.

To better understand grammatical issues that may arise, SMEs need to be able to select from the available dynamic inserts as well as see the set of text items that may be entered for inserts in the message. The form needs to be able to display this information, because it is rather cumbersome for users to keep track of it. This requires adding a new table of dynamic inserts.

6.2 Voice and Language Development Issues

The SMEs who visited from Offutt AFB in August and September, 2010 stressed the importance of the synthetic speech quality for their application. They were concerned about the possibility of creating negative training by using speech that is noticeably synthetic or deviating from the target in dialect or other qualities. Linguists already receive a great deal of training, so any additional training must have compelling benefit. Therefore, attention must be paid to creating synthetic speech of the highest quality possible and as close to the target as possible.

Acoustic Quality of Speech Sounds

The HTS synthetic voices, while being generally good and conveying an impression of speaker individuality, have an assortment of noise artifacts due to local faults in the modeling. The quality could be improved by manually tuning the acoustic parameters to remove acoustic artifacts. Because the process of creating HTS voices includes an irreversible step that produces binary output, this would require building a tool to read the binary files, store the data in ASCII format, and write the data back to binary format.

The quality of the dictionary is crucial to the quality of voices built with HTK/HTS. The dictionaries purchased from XLingual were intended to be used for speech recognition systems, which typically do not use information about the stress level of vowels in their statistical voice models. However, when producing synthetic speech output, vowel stress, for those languages which have a stress distinction among vowels, must be reproduced for natural sounding speech. In addition, the GlobalPhone dictionary transcriptions appeared to be derived by rule, resulting in dubious entries for the many loan words in the corpus, which often do not have regular pronunciations. To get good sound quality using HMM or other machine learning techniques, additional work is required to improve the transcriptions in purchased dictionaries. This work requires input from native speakers to correct or verify transcriptions.

Users in some cases have specific requirements for language dialect. However, there may not be commercially available corpora for the required dialect or dialects. This would necessitate recording native speakers in a sound booth reading appropriate materials to generate sufficient speech for voice modeling. Dialect requirements need to be carefully discussed with the customer before choosing a corpus or speakers to be recorded.

6.3 Emotion Levels

From the beginning of this program, starting with TO 0017 and TO 0007, the synthesis of emotion levels in speech was considered to be a critical requirement. The following are factors that may distinguish different levels of emotion in speech:

- Word choice
- Speech rate
- Loudness
- Pitch level and pitch range
- Voice quality
- Carefulness of articulation
- Errors

Machine learning synthesis techniques such as HTK/HTS cannot capture emotion levels without appropriate audio input that displays the desired emotion levels. Even if such corpora were available, a wholly different model would have to be produced and stored for each emotion level and voice (Schröder, 2001). A more productive route would be to apply rule-based alterations to speech rate, loudness, pitch, and voice quality to derive emotional speech from calm speech using a formant synthesizer (Burkhardt & Sendlmeier, 2000). Some of these alterations can be done with HTS voices; however, voice quality changes are probably not feasible. Formant synthesizers are based on a source/filter model of speech which would allow direct manipulation of the voice source to produce a change in voice quality without affecting the phonemes.

Prosody

Prosody refers to pitch and duration characteristics of speech at the phrase and sentence level. In general, synthetic speech lacks natural prosody. Many synthesis systems do not attempt to model prosody or model it inadequately. Additionally, prosody varies with meaning or speaker intent, so there is no direct mapping between text and prosody. The same phrase may be delivered with a variety of prosodies. So even where there is an adequate prosodic model, there is no way to apply the model to a given utterance without an "understanding" of the context. Applying the model can result in a prosody that is natural and correct for some context, but not correct for the current context.

Unnatural prosody is one of the factors that listeners, such as our SMEs, recognize as deviant in synthetic speech. One researcher studying Computer-Assisted Language Learning (CALL) applications states, "In order to fully meet the requirements of CALL, further attention needs to be paid to accuracy and naturalness, in particular at the prosodic level, and expressiveness" (Handley, 2009).

In our voice models, duration and pitch parameters depend only on the sound itself and its neighbor on either side. This is not a realistic model of how duration and pitch vary in language and produces an unnatural prosody. There are several approaches that could be taken to improve prosody. The voices could be rebuilt using the HTS procedure of context dependent modeling. This would produce more natural speech by using global contextual factors to determine pitch and duration, but the intonation may not be appropriate to our target and would not provide the ability to vary it.

Alternatively, prosody model could be developed, or an open source model could be used to override the pitch and duration values stored in the model. This approach would allow alteration of prosody, for example to produce emphatic and normal versions of the same message in different contexts. This would also provide at least some of the tools needed to produce emotion levels.

6.4 Maintainability of RFTE

As currently written, a set of complex conditions in the XCITE code determine when and how a message is transmitted. It would be desirable to document the conditions with the following goals:

- Provide a general description of factors to consider when adding messaging capability to simulation software.
- Understand the requirements for adding new code to XCITE to trigger messages.

- Convert special purpose code in XCITE for each trigger situation to general purpose data-driven code, where data associated with the message in the database drives decision points in the code, rather than writing special code for that trigger point.

The ultimate goal would be to produce a version of XCITE that allows adding messages with no new coding, but simply adding data to the database. This requires partitioning information between that belonging to code (understood by software developers) and that belonging to the message (understood by SME data providers).

Message Builder and Dynamic Inserts

We have discovered an issue in the message builder code in msg_builder.c. RFTE messages may contain placeholder text called dynamic inserts for which XCITE supplies appropriate content at runtime. For example, “#i” will be replaced with the player’s relative heading. Currently, each dynamic insert can only be used in certain messages. The code that sets the replacement text for the insert is tied to a case statement that specifies the message list for which the insertion takes place. If a SME used a dynamic insert in a message of a different type, the inserted text would not have been initialized, so that an incorrect value or no value would be inserted. The code should be rewritten to set the insertion text variable whether or not it is actually needed, or, alternatively, to set it at the time it is used. This may involve some code reorganization, but it will simplify the code considerably and make the system less fragile by removing an unnecessary linkage between the code and the data.

Voice Server

The Voice Server processes input text through a series of filter-like code functions, handling character case and punctuation, grammar, numbers, letter to phoneme conversion, and phoneme to triphone conversion, before converting the triphone string to audio. The processing is strictly left to right at each level and uses information available only at the current level to produce its output. For the most part, this works well for the current application. However, mixed alphanumeric strings present problems. To handle strings like “MiG-29s”, “hog2”, and “rs23t2@”, where the ‘@’ symbol specifies character-by-character pronunciation, the following sequence is required:

1. Look-up string in the dictionary.
2. If not found, search ahead for ‘@’.
3. If ‘@’ is found, split the string into individual characters.
4. Else, split the string into alpha and numeric sequences (“hog” and “2”).
5. Look up parts in the dictionary.

Doing the look-ahead and multiple dictionary look-ups requires some code rework.

6.5 Centralized *vs.* Decentralized Command and Control

The command and control (C^2) methodology used in XCITE is best described as autonomous or decentralized. Controllers direct actions at the macro level and subordinates execute complex tasks without further direction. This methodology is commonly used in US and North Atlantic. Treaty Organization military communications, but is rarely used by other military organizations. Instead, the military structure of most countries is primarily centralized. Commanders exercise tight, precise, and continuous control over subordinates during military activities. This results in messages being triggered in a manner that is unrealistic for players whose military organizations use centralized C^2, where pilots are under close and continuing direction from airborne or ground controllers and display no autonomy whatsoever in their actions.

All of the foreign languages we modeled use this centralized pattern of command and control communication. XCITE tactic implementation and message triggering requires significant changes to address this methodology. Although we can currently, to some degree, mimic the required behavior, implementing the correct behavior in XCITE is necessary to provide a positive training environment in RFTE and value beyond the RFTE project. Centralized C^2 can be modeled but this will require significant new capability. This will require many more messages as well as more explicit and direct messages and responses during the course of a tactic. Such an implementation would provide more realistic behavior of enemy forces.

6.6 Train for Language Errors

It may be worthwhile to implement language errors of various types into simulations geared toward advanced students. According to SME sources, linguists will encounter errors in the field and need to correctly read the situation in spite of these anomalous utterances. This would require the ability to script messaging, which is not currently implemented in XCITE.

6.7 Further Additions to Message Database

These areas are covered in spreadsheets, but not included in the final message database:

- New material for basic takeoff and landing (TOL) including preflight messages of various kinds for takeoffs and guided approach messages for landings.
- Extended TOL for various aircraft
- Reaction to any target
- Ground training
- Platform specific messages
- Local area navigation and missions in zone
- Long range navigation
- Formation flight
- Bombing
- Air to ground
- Weather
- Scrambled takeoff
- Transfer of aircraft control from one controller to another
- Return to base

7.0 ACCEPTANCE TEST

The acceptance test for MIISTT took place on November 10, 2010. The following were in attendance:

Government:

- Lt Chin Ki Tam
- Larry Boyce

L3/IS:

- Jim Ahern
- Ed Vogel

Contractors:

- Michelle Caisse
- Ronnie Silber

- Raymond Tillman

There were two phases to the acceptance test. In the first phase, conducted in a secure room containing the 32 channel Digital to Analog Convertor (DAC) board and SEGUE emulator software, Test 2 and Test 3 were performed as described in the Acceptance Test Plan (Tillman & Caisse, 2010).

During the demonstration, Jim Ahern, the Systems Engineer for L3/IS, requested tests of issues found in current version of the system installed at Greenville, TX. Their in-house testing focused on recovery from failures of any component of the system.

He requested creating a snapshot of the scenario and reloading it to note the elapsed time displayed in the ThreatIOS interface. In the customer's version of the software, elapsed time, restarted at 0, but in the demonstrated version it correctly retained the saved time. On the other hand, a saved scenario should reload and start with elapsed time of 0. The demonstrated version revealed the same bug as Mr. Ahern had previously encountered: a reloaded scenario retains the elapsed time it had when saved.

Test 3 presented the components of the latest release for inspection: the latest release of XCITE, the newest voice server, and the new voices, as well as the previously delivered voices.

Test 1 was conducted in the Sensitive Compartmented Information Facility (SCIF). It demonstrated the new voices and the capability to run multiple languages and voices within a single scenario. During this part of the acceptance test an altitude value was not output where expected, due to an incorrect symbol in the stored message.

The government representative from Big Safari, Larry Boyce, made positive comments concerning the new delivery. He was pleased with the additional voices and with the synthesis quality. He felt that the system provides a better training tool than anything currently available and one which would undergo continued improvement.

At a meeting following the acceptance test demonstration, it was decided that the software delivery would wait for the release of XCITE 6.0, anticipated for early January 2011.

8.0 SOFTWARE INVENTORY

The following are the products documented by Communications and Computer Systems Requirements Document (CSRD):

Date	Title	Link	License
11/2006	FLITE	http://www.speech.cs.cmu.edu/flite/download.html	redistributable with copyright notice and marked modifications
07/03/2007	OGI toolkit, OGISable, OGIresLPC Festival modules	http://cslu.cse.ogi.edu/tts/download/	registration required
05/072007	Cepstral voices	http://www.cepstral.com/downloads/	free trial

Date	Title	Link	License
12/07/2007	HTS	http://hts.sp.nitech.ac.jp/	simplified BSD
12/07/2007	Hidden Markov Model Toolkit (HTK)	http://htk.eng.cam.ac.uk/	registration required; may not be redistributed
12/07/2007	Speech Signal Processing Toolkit	http://sp-tk.sourceforge.net/	simplified BSD
02/19/2008	DISA Gold	http://iase.disa.mil/stigs/SRR/gdv2-cd1-engine-09/28-2007.iso	
03/07/2008	RedHat 5.0		2 purchased
03/17/2008	Tactical Language and Culture Training Systems	http://www.alelo.com/tactical_language.html	free with registration for U.S. Armed Forces
09/03/2008	XLingual GlobalPhone corpora and dictionaries	http://www-2.cs.cmu.edu/~tanja/GlobalPhone/index-e-wel.html	2 licenses purchased for L-3/Link and 645th.
09/03/2008	LDC CALL Home Lexicon	http://www.ldc.upenn.edu/	2 licenses purchased for L-3/Link and 645th.
02/11/2009	Wall Street Journal CSR-I and CSR-II Sennheiser English audio data set	http://www.ldc.upenn.edu/	2 licenses purchased for L-3/Link and 645th.
no CSRD	Foreign Language CDs from DLI	http://www.dliflc.edu/languageresources.html	
no CSRD	NOISE-ROM-0 (1 CD)	Institute for Perception-TNO, The Netherlands	unrestricted
no CSRD	NOISEX-92 (2 CDs)	Institute for Perception-TNO, The Netherlands	unrestricted

8.1 Licensing Restrictions on Voice Corpora and Derived Works

Voices were created using source data acquired from a CMU research group, Linguistic Data Consortium, and XLingual GlobalPhone. The Linguistic Data Consortium and XLingual provide a significant number of linguistic corpora that were used to create word to phoneme dictionaries, construct voices, and assist in building letter to sound rules. This is copyrighted material and requires user licenses. It is unclear whether voices created from these corpora are considered “derived” products and thus non-transferrable without permission of the provider. It is unclear whether the general purpose dictionaries created from these corpora would be considered “derived” with the same result. L-3 Link prepared and submitted a request for a legal opinion regarding these issues, but they declined to express an opinion regarding government-provided data. L-3 Link has turned the licensing issue for government-provided data over to the government to handle (Hoefer, 3 June 2008).

9.0 ACRONYMS

711 HPW/RHA	711th Human Performance Wing, Human Effectiveness Directorate, Warfighter Readiness Research Division (Mesa, AZ)
711 HPW/RHX	711th Human Performance Wing, Human Effectiveness Directorate, Anticipate & Influence Behavior Division (Dayton, OH)
AFB	Air Force Base
AFRL	Air Force Research Laboratory
C2	Command and Control
C4ISR	Command, Control, Communications, Computers, Intelligence, Surveillance and Reconnaissance
CALL	Computer-Assisted Language Learning
CGF	Computer Generated Force
CNI	Communications, Navigation, and Identification
CSRD	Communications-Computer Systems Requirements Document
CVS	Concurrent Versions Source
DAC	Digital to Analog Convertor
DMO	Distributed Mission Operations
HMM	Hidden Markov Model
HTK	Hidden Markov Model Toolkit
HTS	HMM-based Speech Synthesis System
IOS	Instructor Operator Station
ISR	Intelligence, Surveillance, and Reconnaissance
LDC	Linguistic Data Consortium
LDOM	Limited Domain
MIISTT	Multi-INT and Information Operations Simulation and Training Technologies
NATO	North Atlantic Treaty Organization
NSA	National Security Agency
PI	Principle Investigator
RF	Radio Frequency
RFTE	Radio Frequency Threat Environment
RJTS	Rivet Joint Training System
SCIF	Sensitive Compartmented Information Facility
SEGUE	Simulation Environment Generation User Interface
SME	Subject Matter Expert
SOW	Statement of Work
TOL	Takeoff and Landing
TTS	Text-to-Speech
XCITE	eXpert Common Immersive Theater Environment

10.0 DOCUMENTS SUPPORTING TASK ORDER 0026

Burlant, J. G. (June 2007). *Radio Frequency Threat Environment (RFTE) audio database and software development*. Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Kaplan, H., & Linford, M. (June 2006). *Cryptologic linguist training: future language-training technology concepts*. Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Science Applications International Corporation (SAIC). (30 December 2008). *ISRD White Cell Support (ISRD)* - *Technical Report – Feasibility Study* (unpublished report to L-3 Communications). Beavercreek, OH: Author.

Science Applications International Corporation (SAIC). (30 July 2008). *ISRD White Cell* - *Version Description Document* (VDD) v1.0 (unpublished report to L-3 Communications). Beavercreek, OH: Author.

Silber, R. F., & Burlant, J. G. (3 January 2007). *Radio Frequency Threat Environment (RFTE) multilanguage trainer technologies analysis* (AFRL-HE-AZ-TR-2007-0039;ADB335143). Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Tillman, R. T., & Caisse, M. (15 September 2010). *Multi-INT and Information Operations Simulation and Training Technologies (MIISTT) Acceptance test plan*. Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Tygret, K. I., Lewandowski, D. H., & Jones, T. R. (April 2007). *Expert Common Immersive Theater Environment – Research & Development (XCITE $^{R\&D}$)* Version 1.0 (Software Users Manual). Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Tygret, K. I., Silber, R. F., Burlant, J. G., & Hoefer, S. J. (August, 2005). *Final Report MTL RFT Multilanguage Trainer*. Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Woudstra, D. (13 January 2009). *Radio Frequency Threat Environment (RFTE) Engineering Development, Integration & Delivery Support (Engineering Procedures Manual).* Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

10.1 Documents for Commercial Customer L-3 Communications Integrated Systems

Hoefer, S. J. (3 June 2008). *RFTE (Radio Frequency Threat Environment) Critical Design Review (CDR).* PowerPoint presentation. Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Hoefer, S. J. (15 June 2008). *Radio Frequency Threat Environment (RFTE) Engineering Development, Integration & Delivery Support* (Interface Description Document). Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Hoefer, S. J. (4 June 2009). *Radio Frequency Threat Environment (RFTE) Engineering Development, Integration & Delivery Support* (Engineering Procedures Manual: Integration/Troubleshooting Guide). Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Hoefer, S. J. (27 March 2008). *Radio Frequency Threat Environment (RFTE) System Requirements Specification (SRS)*. Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Hoefer, S. J. (23 April 2008). *Radio Frequency Threat Environment (RFTE) Engineering Development, Integration & Delivery Support* (Software Design Document), Rev N/C. Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Hoefer, S. J. (28 March 2008). *Radio Frequency Threat Environment (RFTE) Engineering Development, Integration & Delivery Support* (Interface Control Document), Rev A. Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Hoefer, S. J. (28 March 2008). *Radio Frequency Threat Environment (RFTE) Engineering Development, Integration & Delivery Support* (Interface Description Document), Rev N/C8. Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Hoefer, S. J. (25 March 2008). *Radio Frequency Threat Environment (RFTE) Engineering Development, Integration & Delivery Support* (System Requirements Specification), Rev B. Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Hoefer, S. J. (23 January 2008). *Radio Frequency Threat Environment (RFTE) Engineering Development, Integration & Delivery Support* (Software Requirements Specification), Rev N/C. Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Tillman, R. T. (18 January 2008). *Radio Frequency Threat Environment (RFTE) Engineering Development, Integration & Delivery Support* (Integration and Test Plan), Rev N/C. Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Tillman, R. T. (09 September 2009). *Radio Frequency Threat Environment (RFTE) Engineering Development, Integration & Delivery Support* (Voice Server Application Users Guide). Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Wight, D. R. (23 April 2008). *Radio Frequency Threat Environment (RFTE) Engineering Development, Integration & Delivery Support (Software Design Document).* Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

11.0 REFERENCES CITED

Burkhardt, F., & Sendlmeier, W. F. (2000). Verification of acoustical correlates of emotional speech using formant-synthesis. In *Proceedings of the ISCA Workshop on Speech and Emotion* (pp. 151-156). Belfast, Northern Ireland.

Handley, Z. (October 2009). Is text-to-speech synthesis ready for use in computer-assisted language learning? *Speech Communication 51*(10), 906-919.

NOISE-ROM-0, NATO: AC243/(Panel 3)/RSG-10 ESPRIT: Project No. 2589-SAM. Produced by: Institute for Perception-TNO, The Netherlands, Speech Research Unit, RSRE, United Kingdom. Copyright: TNO, Soesterberg, The Netherlands, Feb 1990. For more information: Institute for Perception-TNO, PO-box 23, 3769 ZG Soesterberg, The Netherlands.

NOISEX-92. Contact The Speech Research Unit, Ex1, DRA Malvern, St.Andrew's Road, Malvern, Worcestershire, WR14 3PS, UK. Tel +44-684-894074 Fax +44-684-894384, or see http://spib.rice.edu/spib/select_noise.html for a subset of the database.

Powers, R. (n.d.). *Air Force enlisted job descriptions: 1A8X1 - Airborne Cryptologic Linguist.* Retrieved from About.com http://usmilitary.about.com/od/airforceenlistedjobs/a/afjob1a8x1.htm.

Schröder, M. (2001). Emotional Speech Synthesis - A Review. In *Proceedings of the 7th European Conference on Speech Communication and Technology* (EUROSPEECH), (Vol. 1, pp. 561-564). Aalborg, Denmark.

12.0 RECOMMENDED BACKGROUND READING

Black, A. W. (2003). Unit Selection and Emotional Speech. In *Proceedings of EUROSPEECH 2003*, Geneva, Switzerland.

Black, A. W., & Lenzo, K. (2000). *Building Voices in the Festival Speech Synthesis System*, DRAFT (updated 2003).

Black, A. W., & Lenzo, K. (2000). Limited Domain Synthesis. In *Proceedings of ICSLP2000*, Beijing, China.

Black, A. W., & Lenzo, K. (2001). Flite: a small fast run-time synthesis engine. In *Proceedings of ISCA, 4th Speech Synthesis Workshop* (pp 157-162), Scotland.

Black, A. W., Taylor, P., & Caley, R. (27 December 2002). *The Festival Speech Synthesis system: System documentation*, Ed. 1.4, for Festival Version 1.4.3. http://festvox.org/docs/manual-1.4.3/festival_toc.html.

Black, A. W., Zen, H., & Tokuda, K. (April 2007). Statistical parametric speech synthesis. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP) (pp. 1229-1232).

Hemptinne, C. (June 2006). *Integration of the harmonic plus noise model (HNM) into the hidden Markov model-based speech synthesis system (HTS)* (Master thesis). IDIAP Research Institute, Valais, Switzerland.

Kim, S.-J., Kim, J.-J., & Hahn, M.-S. (2006). Implementation and evaluation of an HMM-based Korean speech synthesis system. *IEICE - Transactions on Information and Systems*, *E89-D*(3), 1116-1119.

Kishore, S., & Black, A. W. (2003). Unit Size in Unit Selection Speech Synthesis, In *Proceedings of EUROSPEECH 2003*, Geneva, Switzerland.

Kominek, J., & Black, A. W. (2003). *CMU ARCTIC databases for speech synthesis* (Tech Report CMU-LTI-03-177). Pittsburg, PA: Language Technologies Institute, School of Computer Science, Carnegie Mellon University.

Shriberg, E., Bates, R., Taylor, P., Stolcke, A., Ries, K., Jurafsky, D., Coccaro, N., Martin, R., Meteer, M., & Van Ess -Dykema, C. (1998). Can prosody aid the automatic classification of dialog acts in conversational speech? *Language and Speech*, *41,* 3-4.

Stolcke, A., Coccaro, N., Bates, R., Taylor, P., Van Ess-Dykema, C., Ries, K., Shriberg, E., Jurafsky, D., Martin, R., & Meteer, M. (2000). Dialog act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, *26*(3), 339–373.

Tachibana, M., Yamagishi, J., Masuko, T., & Kobayashi, T. (November 2005). Speech synthesis with various emotional expressions and speaking styles by style Interpolation and morphing. *IEICE - Transactions on Information and Systems, E88-D*(11), 2484-2491.

Taylor, P. (2000). Analysis and synthesis of intonation using the tilt model. *Journal of the Acoustical Society of America, 107*(3):1697-1714.

Taylor, P., & Black, A. W. (1998). Assigning phrase breaks from part of speech sequences. *Computer Speech and Language, 12,* 99-117.

Taylor, P., Black, A. W., & Caley, R. (1998). The architecture of the Festival Speech Synthesis System. In *3rd ESCA Workshop on Speech Synthesis* (pp. 147-151), Jenolan Caves, Australia,

Taylor, P., Caley, R., Black, A. W., & King, S. (15 June 1999). *Edinburgh Speech Tools Library: System Documentation* Ed. 1.2, for 1.2.0. http://festvox.org/docs/speech_tools-1.2.0/book1.htm.

Tokuda, K., Zen, H., & Black, A. W. (2004). HMM-based approach to multilingual speech synthesis. In S. Narayanan, & A. Alwan (Eds.) *Text to speech synthesis: New paradigms and advances* (pp. 135-153). Upper Saddle River, NJ: Prentice Hall.

Tokuda, K., Zen, H., & Black, A. W. (September 2002). An HMM-based speech synthesis system applied to English. In *Proceedings of 2002 IEEE International Workshop on Software Stability at Work* (pp. 11-13).

Tokuda, K., Mausko, T., Miyazaki, N., & Kobayashi, T. (March 2002). Multi-space probability distribution HMM. *IEICE TRANSACTIONS on Information and Systems, E85-D*(3), 455-464.

Tokuda, K., Yoshimura, T., Masuko, T., Kobayashi, T., & Kitamura, T. (June 2000). Speech parameter generation algorithms for HMM-based speech synthesis. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP '00), vol. 3, (pp. 1315–1318), Istanbul, Turkey.

Tomokiyo, L. M., Black, A. W., & Lenzo, K. (2003). Arabic in my Hand: Small-footprint synthesis of Egyptian Arabic. In *Proceedings of EUROSPEECH 2003*, Geneva, Switzerland (pp. 2049-2052).

Yamagishi, J., Onishi, K., Masuko, T., &Kobayashi, T. (March 2005). Acoustic modeling of speaking styles and emotional expressions in HMM-based speech synthesis. *IEICE - Transactions on Information and Systems, E88-D*(3), 503-509.

Yoshimura, T. (January 2002). *Simultaneous modeling of phonetic and prosodic parameters, and characteristic conversion for HMM-based text-to-speech systems* (Ph.D thesis). Nagoya Institute of Technology, Nagoya, Japan.

Yoshimura, T., Tokuda, K., Masuko, T., Kobayashi, T., & Kitamura, T. (Sept. 1999). Simultaneous modeling of spectrum, pitch and duration in HMM-based speech synthesis. In *Proceedings of EUROSPEECH1999* (pp. 2347-2350).

Zen, H., Nose, T., Yamagishi, J., Sako, S., Masuko, T., Black, A. W., & Tokuda, K. (August 2007). The HMM-based speech synthesis system version 2.0. In *6th ISCA Workshop on Speech Synthesis*, Bonn, Germany, Bonn, Germany.

Zen, H., Tokuda, K., & Kitamura, T. (October 2004). An introduction of trajectory model into HMM-based speech synthesis. In *Proceedings of 5th ISCA Speech Synthesis Workshop* (pp. 191-196), June 2004.

Zen, H., Tokuda, K., Masuko, T., Kobayashi, T., &Kitamura, T. (2004). Hidden semi-Markov model based speech synthesis. In *Proceedings of ICSLP 2004*, vol. II, (pp. 1397-1400).